# Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification Rather Than Natural Evolution and Delineation of Its Probable Synthetic Route

Li-Meng Yan (MD, PhD)[1], Shu Kang (PhD)[1], Shanchang Hu (PhD)[1]

[1]Yan Research – An Independent Research Team

Correspondence: team.lmyan@gmail.com

**Abstract**

The COVID-19 pandemic caused by the novel coronavirus SARS-CoV-2 has led to over 910,000 deaths worldwide and unprecedented decimation of the global economy. Despite its tremendous impact, the origin of SARS-CoV-2 has remained mysterious and controversial. The natural origin theory, although widely accepted, lacks substantial support. The alternative theory that the virus may have come from a research laboratory is, however, strictly censored on peer-reviewed scientific journals. Nonetheless, SARS-CoV-2 shows biological characteristics that are inconsistent with a naturally occurring, zoonotic virus. In this report, we describe the genomic, structural, medical, and literature evidence, which, when considered together, strongly contradicts the natural origin theory. The evidence shows that SARS-CoV-2 should be a laboratory product created by using bat coronaviruses ZC45 and/or ZXC21 as a template and/or backbone. Building upon the evidence, we further postulate a synthetic route for SARS-CoV-2, demonstrating that the laboratory-creation of this coronavirus is convenient and can be accomplished in approximately six months. Our work emphasizes the need for an independent investigation into the relevant research laboratories. It also argues for a critical look into certain recently published data, which, albeit problematic, was used to support and claim a natural origin of SARS-CoV-2. From a public health perspective, these actions are necessary as knowledge of the origin of SARS-CoV-2 and of how the virus entered the human population are of pivotal importance in the fundamental control of the COVID-19 pandemic as well as in preventing similar, future pandemics.

**Publication Note (July 17th, 2021):**

The three Yan reports used scientific evidence and analyses to prove that SARS-CoV-2 is an *Unrestricted Bioweapon* created by military scientists of the Chinese Communist Party (CCP) regime. These reports have played a pivotal role in revealing the true identity of the ongoing *Unrestricted Biowarfare*. For this reason, the CCP and its allies have been constantly launching attacks at the Yan Reports. Very recently, the *Rule of Law Foundation* (ROLF) and *Rule of Law Society* (ROLS), which we have listed as our honorary affiliation in our reports, requested *Zenodo* to have the original uploads of our reports closed. This was done by the ROLF & ROLS without informing us authors or seeking our agreement. This is unacceptable because the work was done by us authors independently with no financial assistance provided by the ROLF & ROLS or any other organization. Their action here has no scientific

basis and is against the rules of scientific publications. To restore the availability of our reports to the world, we have therefore re-uploaded the three Yan reports. Our affiliation has been changed to *Yan Research – An Independent Research Team*.

The current report was originally published on September 14th, 2020. As of July 16th, 2021, the original *Zenodo* upload of it has been viewed 1,339,786 times and downloaded 797,325 times. Upon mutual agreement, Dr. Jie Guan opted out of this publication and his contributions have instead been specified in the acknowledgements.

## Introduction

COVID-19 has caused a world-wide pandemic, the scale and severity of which are unprecedented. Despite the tremendous efforts taken by the global community, management and control of this pandemic remains difficult and challenging.

As a coronavirus, SARS-CoV-2 differs significantly from other respiratory and/or zoonotic viruses: it attacks multiple organs; it is capable of undergoing a long period of asymptomatic infection; it is highly transmissible and significantly lethal in high-risk populations; it is well-adapted to humans since the very start of its emergence[1]; it is highly efficient in binding the human ACE2 receptor (hACE2), the affinity of which is greater than that associated with the ACE2 of any other potential host[2,3].

The origin of SARS-CoV-2 is still the subject of much debate. A widely cited *Nature Medicine* publication has claimed that SARS-CoV-2 most likely came from nature[4]. However, the article and its central conclusion are now being challenged by scientists from all over the world[5-15]. In addition, authors of this *Nature Medicine* article show signs of conflict of interests[16,17], raising further concerns on the credibility of this publication.

The existing scientific publications supporting a natural origin theory rely heavily on a single piece of evidence – a previously discovered bat coronavirus named RaTG13, which shares a 96% nucleotide sequence identity with SARS-CoV-2[18]. However, the existence of RaTG13 in nature and the truthfulness of its reported sequence are being widely questioned[6-9,19-21]. It is noteworthy that scientific journals have clearly censored any dissenting opinions that suggest a non-natural origin of SARS-CoV-2[8,22]. Because of this censorship, articles questioning either the natural origin of SARS-CoV-2 or the actual existence of RaTG13, although of high quality scientifically, can only exist as preprints[5-9,19-21] or other non-peer-reviewed articles published on various online platforms[10-13,23]. Nonetheless, analyses of these reports have repeatedly pointed to severe problems and a probable fraud associated with the reporting of RaTG13[6,8,9,19-21]. Therefore, the theory that fabricated scientific data has been published to mislead the world's efforts in tracing the origin of SARS-CoV-2 has become substantially convincing and is interlocked with the notion that SARS-CoV-2 is of a non-natural origin.

Consistent with this notion, genomic, structural, and literature evidence also suggest a non-natural origin of SARS-CoV-2. In addition, abundant literature indicates that gain-of-function research has long advanced to the stage where viral genomes can be precisely engineered and manipulated to enable the creation of novel coronaviruses possessing unique properties. In this report, we present such evidence and the associated analyses. Part 1 of the report describes the genomic and structural features of SARS-CoV-2, the presence of which could be consistent with the theory that the virus is a product of laboratory modification beyond what could be afforded by simple serial viral passage. Part 2 of the report describes a highly probable pathway for the laboratory creation of SARS-CoV-2, key steps of which are supported by evidence present in the viral genome. Importantly, part 2 should be viewed as a demonstration of how SARS-CoV-2 could be conveniently created in a laboratory in a short period of time using available materials and well-documented techniques. This report is produced by a team of experienced scientists using our combined expertise in virology, molecular biology, structural biology, computational biology, vaccine development, and medicine.

## 1. Has SARS-CoV-2 been subjected to *in vitro* manipulation?

We present three lines of evidence to support our contention that laboratory manipulation is part of the history of SARS-CoV-2:

i. The genomic sequence of SARS-CoV-2 is suspiciously similar to that of a bat coronavirus discovered by military laboratories in the Third Military Medical University (Chongqing, China) and the Research Institute for Medicine of Nanjing Command (Nanjing, China).

ii. The receptor-binding motif (RBM) within the Spike protein of SARS-CoV-2, which determines the host specificity of the virus, resembles that of SARS-CoV from the 2003 epidemic in a suspicious manner. Genomic evidence suggests that the RBM has been genetically manipulated.

iii. SARS-CoV-2 contains a unique furin-cleavage site in its Spike protein, which is known to greatly enhance viral infectivity and cell tropism. Yet, this cleavage site is completely absent in this particular class of coronaviruses found in nature. In addition, rare codons associated with this additional sequence suggest the strong possibility that this furin-cleavage site is not the product of natural evolution and could have been inserted into the SARS-CoV-2 genome artificially by techniques other than simple serial passage or multi-strain recombination events inside co-infected tissue cultures or animals.

### 1.1 Genomic sequence analysis reveals that ZC45, or a closely related bat coronavirus, should be the backbone used for the creation of SARS-CoV-2

The structure of the ~30,000 nucleotides-long SARS-CoV-2 genome is shown in Figure 1. Searching the NCBI sequence database reveals that, among all known coronaviruses, there were two related bat coronaviruses, ZC45 and ZXC21, that share the highest sequence identity with SARS-CoV-2 (each bat coronavirus is ~89% identical to SARS-CoV-2 on the nucleotide level). Similarity between the genome of SARS-CoV-2 and those of representative β coronaviruses is depicted in Figure 1. ZXC21, which is 97% identical to and shares a very similar profile with ZC45, is not shown. Note that the RaTG13 virus is excluded from this analysis given the strong evidence suggesting that its sequence may have been fabricated and the virus does not exist in nature[2,6-9]. (*A follow-up report, which summarizes the up-to-date evidence proving the spurious nature of RaTG13, will be submitted soon*)

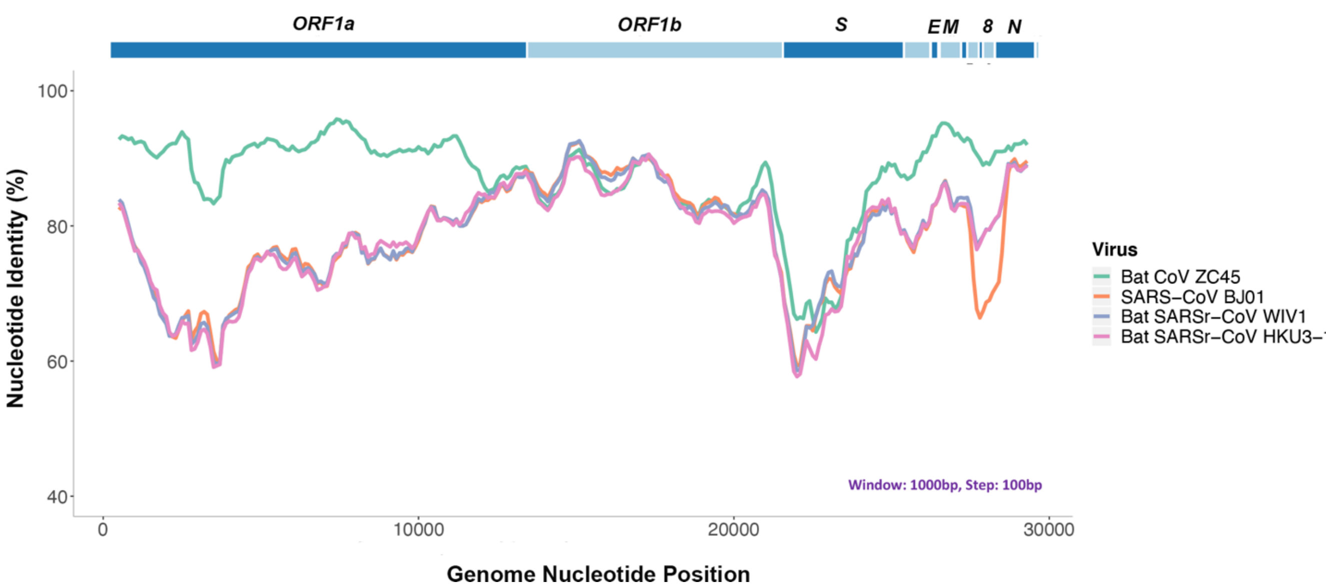


***Figure 1. Genomic sequence analysis reveals that bat coronavirus ZC45 is the closest match to SARS-CoV-2.*** *Top: genomic organization of SARS-CoV-2 (2019-nCoV WIV04). Bottom: similarity plot based on the full-length genome of 2019-nCoV WIV04. Full-length genomes of SARS-CoV BJ01, bat SARSr-CoV WIV1, bat SARSr-CoV HKU3-1, bat coronavirus ZC45 were used as reference sequences.*

When SARS-CoV-2 and ZC45/ZXC21 are compared on the amino acid level, a high sequence identity is observed for most of the proteins. The Nucleocapsid protein is 94% identical. The Membrane protein is 98.6% identical. The S2 portion (2nd half) of the Spike protein is 95% identical. Importantly, the Orf8 protein is 94.2% identical and the E protein is 100% identical.

Orf8 is an accessory protein, the function of which is largely unknown in most coronaviruses, although recent data suggests that Orf8 of SARS-CoV-2 mediates the evasion of host adaptive immunity by downregulating MHC-I[24]. Normally, Orf8 is poorly conserved in coronaviruses[25]. Sequence blast indicates that, while the Orf8 proteins of ZC45/ZXC21 share a 94.2% identity with SARS-CoV-2 Orf8, no other coronaviruses share more than 58% identity with SARS-CoV-2 on this particular protein. The very high homology here on the normally poorly conserved Orf8 protein is highly unusual.

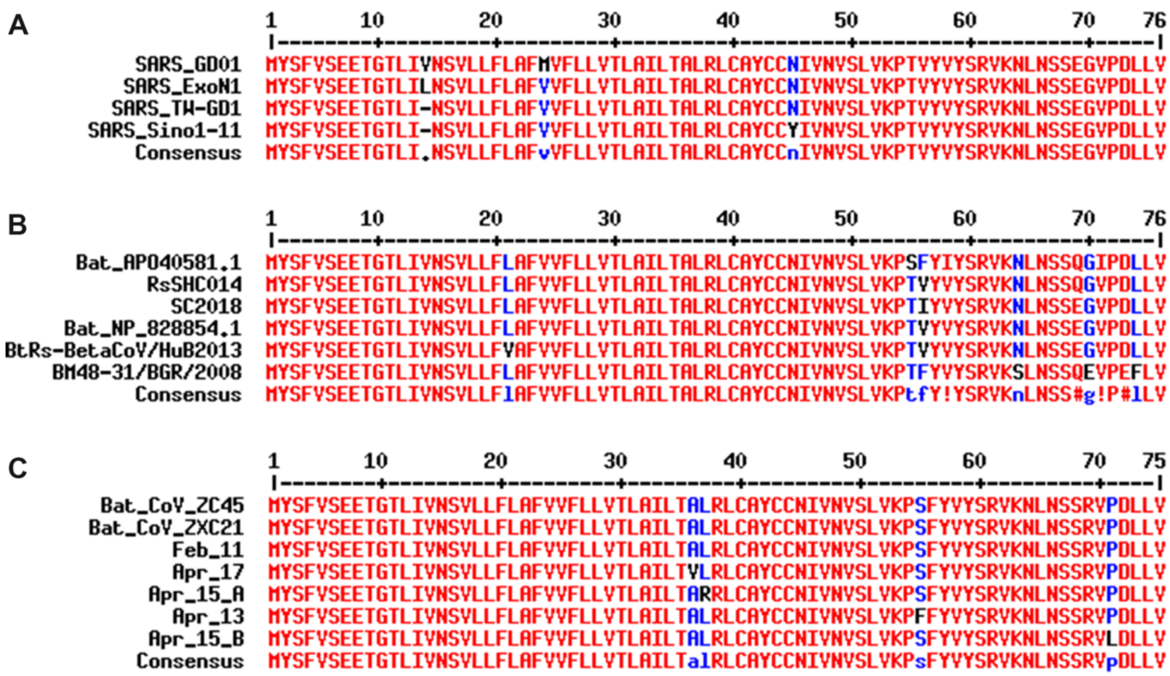



***Figure 2. Sequence alignment of the E proteins from different β coronaviruses demonstrates the E protein's permissiveness and tendency toward amino acid mutations.*** *A. Mutations have been observed in different strains of SARS-CoV. GenBank accession numbers: SARS_GD01: AY278489.2, SARS_ExoN1: ACB69908.1, SARS_TW_GD1: AY451881.1, SARS_Sino1_11: AY485277.1. B. Alignment of E proteins from related bat coronaviruses indicates its tolerance of mutations at multiple positions. GenBank accession numbers: Bat_AP040581.1: APO40581.1, RsSHC014: KC881005.1, SC2018: MK211374.1, Bat_NP_828854.1: NP_828854.1, BtRs-BetaCoV/HuB2013: AIA62312.1, BM48-31/BGR/2008: YP_003858586.1. C. While the early copies of SARS-CoV-2 share 100% identity on the E protein with ZC45 and ZXC21, sequencing data of SARS-CoV-2 from April 2020 indicates that mutation has occurred at multiple positions. Accession numbers of viruses: Feb_11: MN997409, ZC45: MG772933.1, ZXC21: MG772934, Apr_13: MT326139, Apr_15_A: MT263389, Apr_15_B: MT293206, Apr_17: MT350246. Alignments were done using the MultAlin webserver (http://multalin.toulouse.inra.fr/multalin/).*

The coronavirus E protein is a structural protein, which is embedded in and lines the interior of the membrane envelope of the virion[26]. The E protein is tolerant of mutations as evidenced in both SARS (Figure 2A) and related bat coronaviruses (Figure 2B). This tolerance to amino acid mutations of the E protein is further evidenced in the current SARS-CoV-2 pandemic. After only a short two-month spread of the virus since its outbreak in humans, the E proteins in SARS-CoV-2 have already undergone mutational changes. Sequence data obtained during the month of April reveals that mutations have occurred at four different locations in different strains (Figure 2C). Consistent with this finding, sequence blast analysis indicates that, with the exception of SARS-CoV-2, no known coronaviruses share 100% amino acid sequence identity on the E protein with ZC45/ZXC21 (*suspicious coronaviruses published after the start of the current pandemic are excluded*[18,27-31]). Although 100% identity on the E protein has been observed between SARS-CoV and certain SARS-related bat coronaviruses, none of those pairs simultaneously share over 83% identity on the Orf8 protein[32]. Therefore, the 94.2% identity on the Orf8 protein, 100% identity on the E protein, and the overall genomic/amino acid-level resemblance between SARS-CoV-2 and ZC45/ZXC21 are highly unusual. Such evidence, when considered together, is consistent with a hypothesis that the SARS-CoV-2 genome has an origin based on the use of ZC45/ZXC21 as a backbone and/or template for genetic gain-of-function modifications.

Importantly, ZC45 and ZXC21 are bat coronaviruses that were discovered (between July 2015 and February 2017), isolated, and characterized by military research laboratories in the Third Military Medical University (Chongqing, China) and the Research Institute for Medicine of Nanjing Command (Nanjing, China). The data and associated work were published in 2018[33,34]. Clearly, this backbone/template, which is essential for the creation of SARS-CoV-2, exists in these and other related research laboratories.

What strengthens our contention further is the published RaTG13 virus[18], the genomic sequence of which is reportedly 96% identical to that of SARS-CoV-2. While suggesting a natural origin of SARS-CoV-2, the RaTG13 virus also diverted the attention of both the scientific field and the general public away from ZC45/ZXC21[4,18]. In fact, a Chinese BSL-3 lab (the Shanghai Public Health Clinical Centre), which published a *Nature* article reporting a conflicting close phylogenetic relationship between SARS-CoV-2 and ZC45/ZXC21 rather than with RaTG13[35], was quickly shut down for "rectification"[36]. It is believed that the researchers of that laboratory were being punished for having disclosed the SARS-CoV-2—ZC45/ZXC21 connection. On the other hand, substantial evidence has accumulated, pointing to severe problems associated with the reported sequence of RaTG13 as well as questioning the actual existence of this bat virus in nature[6,7,19-21]. A very recent publication also indicated that the receptor-binding domain (RBD) of the RaTG13's Spike protein could not bind ACE2 of two different types of horseshoe bats (they closely relate to the horseshoe bat *R. affinis*, RaTG13's alleged natural host)[2], implicating the inability of RaTG13 to infect horseshoe bats. This finding further substantiates the suspicion that the reported sequence of RaTG13 could have been fabricated as the Spike protein encoded by this sequence does not seem to carry the claimed function. The fact that a virus has been fabricated to shift the attention away from ZC45/ZXC21 speaks for an actual role of ZC45/ZXC21 in the creation of SARS-CoV-2.

### 1.2 The receptor-binding motif of SARS-CoV-2 Spike cannot be born from nature and should have been created through genetic engineering

The Spike proteins decorate the exterior of the coronavirus particles. They play an important role in infection as they mediate the interaction with host cell receptors and thereby help determine the host range and tissue tropism of the virus. The Spike protein is split into two halves (Figure 3). The front or N-terminal half is named S1, which is fully responsible for binding the host receptor. In both SARS-CoV

and SARS-CoV-2 infections, the host cell receptor is hACE2. Within S1, a segment of around 70 amino acids makes direct contacts with hACE2 and is correspondingly named the receptor-binding motif (RBM) (Figure 3C). In SARS-CoV and SARS-CoV-2, the RBM fully determines the interaction with hACE2. The C-terminal half of the Spike protein is named S2. The main function of S2 includes maintaining trimer formation and, upon successive protease cleavages at the S1/S2 junction and a downstream S2' position, mediating membrane fusion to enable cellular entry of the virus.

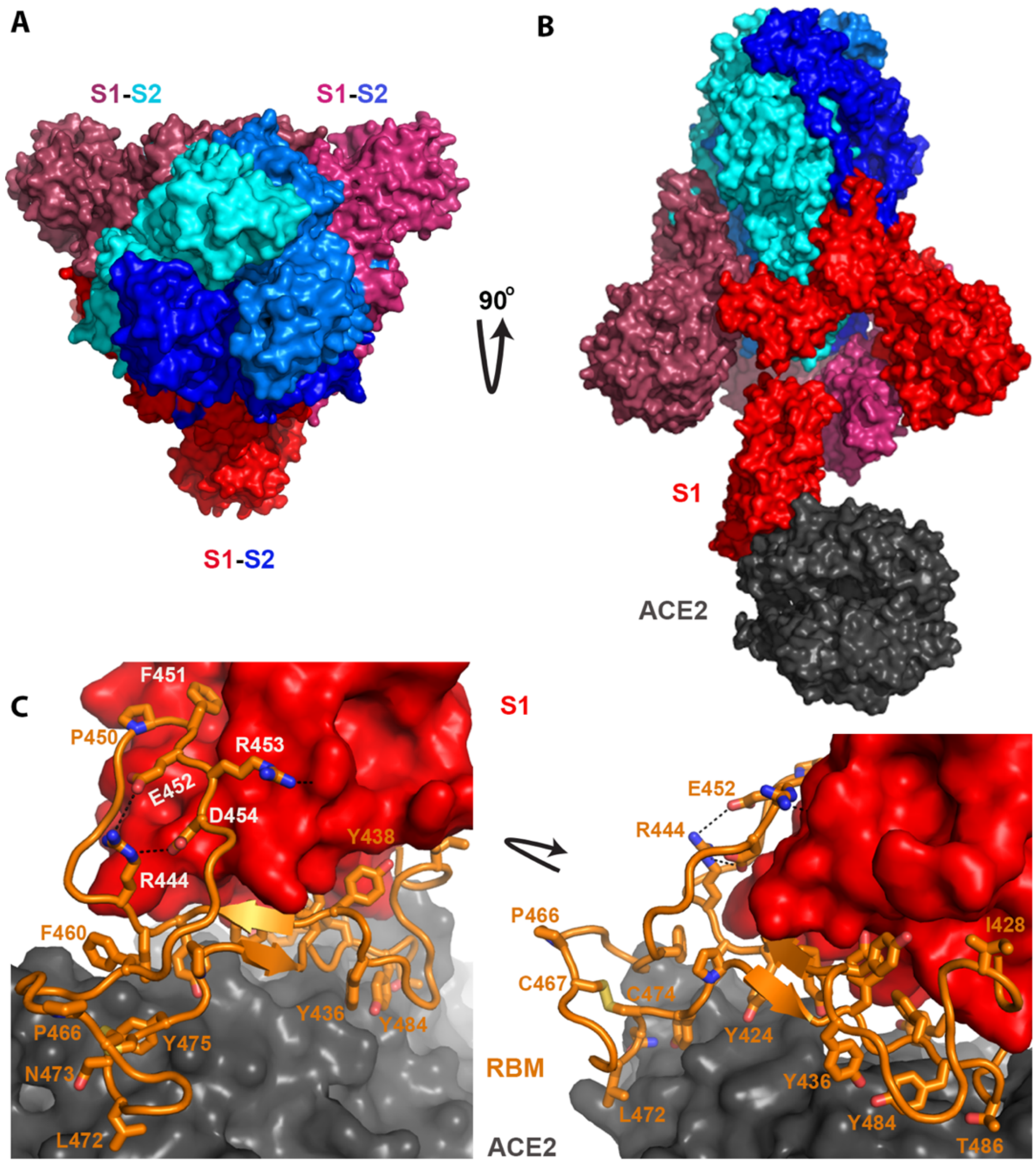



***Figure 3. Structure of the SARS Spike protein and how it binds to the hACE2 receptor.*** *Pictures were generated based on PDB ID: 6acj[37]. A) Three spike proteins, each consisting of a S1 half and a S2 half, form a trimer. B) The S2 halves (shades of blue) are responsible for trimer formation, while the S1 portion (shades of red) is responsible for binding hACE2 (dark gray). C) Details of the binding between S1 and hACE2. The RBM of S1, which is important and sufficient for binding, is colored in orange. Residues within the RBM that are important for either hACE2 interaction or protein folding are shown as sticks (residue numbers follow the SARS Spike sequence).*

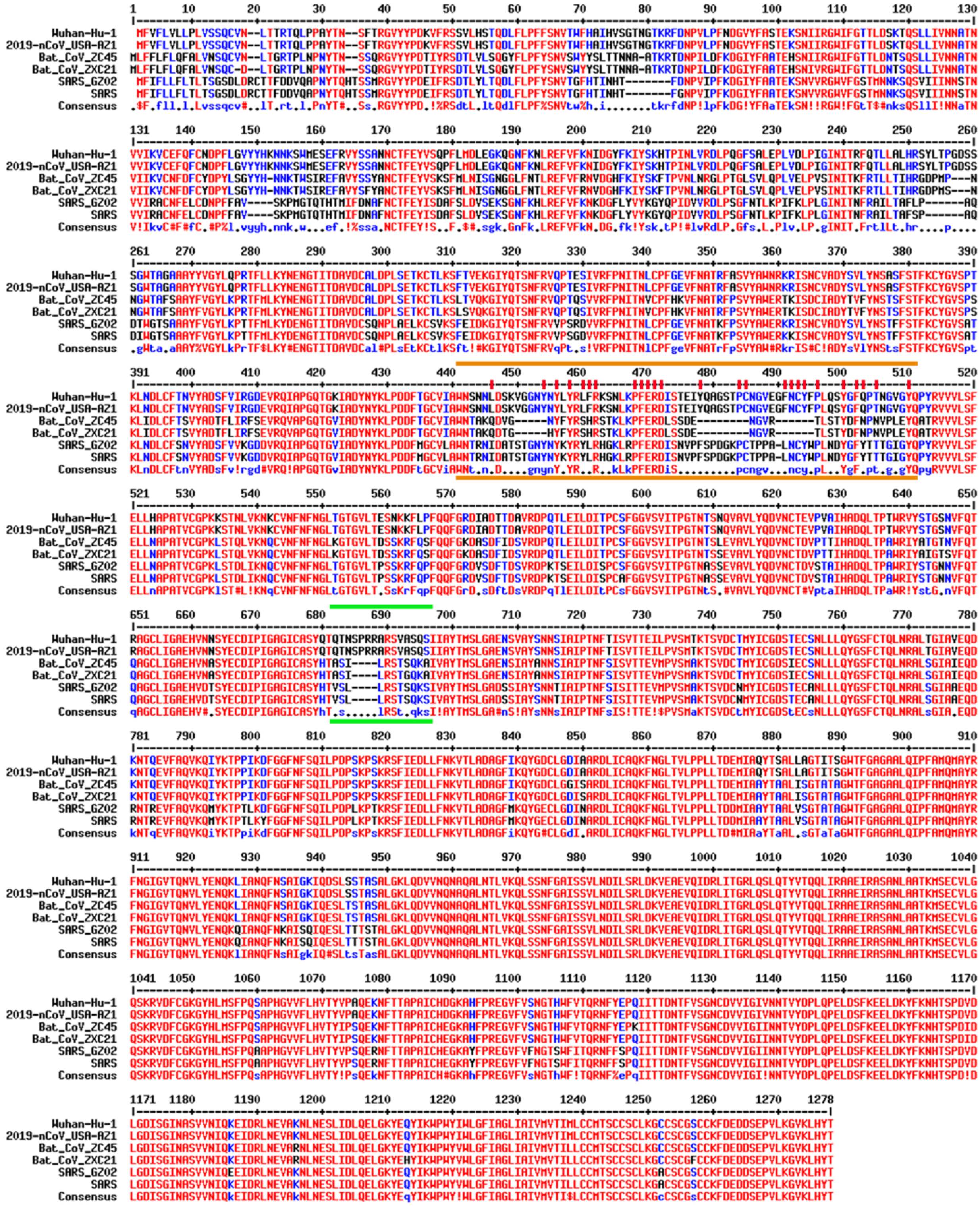


***Figure 4. Sequence alignment of the spike proteins from relevant coronaviruses****.* *Viruses being compared include SARS-CoV-2 (Wuhan-Hu-1: NC_045512, 2019-nCoV_USA-AZ1: MN997409), bat coronaviruses (Bat_CoV_ZC45: MG772933, Bat_CoV_ZXC21: MG772934), and SARS coronaviruses (SARS_GZ02: AY390556, SARS: NC_004718.3). Region marked by two orange lines is the receptor-binding motif (RBM), which is important for interaction with the hACE2 receptor. Essential residues are additionally highlighted by red sticks on top. Region marked by two green lines is a furin-cleavage site that exists only in SARS-CoV-2 but not in any other lineage B β coronavirus.*

Similar to what is observed for other viral proteins, S2 of SARS-CoV-2 shares a high sequence identity (95%) with S2 of ZC45/ZXC21. In stark contrast, between SARS-CoV-2 and ZC45/ZXC21, the S1 protein, which dictates which host (human or bat) the virus can infect, is much less conserved with the amino acid sequence identity being only 69%.

Figure 4 shows the sequence alignment of the Spike proteins from six β coronaviruses. Two are viruses isolated from the current pandemic (Wuhan-Hu-1, 2019-nCoV_USA-AZ1); two are the suspected template viruses (Bat_CoV_ZC45, Bat_CoV_ZXC21); two are SARS coronaviruses (SARS_GZ02, SARS). The RBM is highlighted in between two orange lines. Clearly, despite the high sequence identity for the overall genomes, the RBM of SARS-CoV-2 differs significantly from those of ZC45 and ZXC21. Intriguingly, the RBM of SARS-CoV-2 resembles, on a great deal, the RBM of SARS Spike. Although this is not an exact "copy and paste", careful examination of the Spike-hACE2 structures[37,38] reveals that all residues essential for either hACE2 binding or protein folding (orange sticks in Figure 3C and what is highlighted by red short lines in Figure 4) are "kept". Most of these essential residues are precisely preserved, including those involved in disulfide bond formation (C467, C474) and electrostatic interactions (R444, E452, R453, D454), which are pivotal for the structural integrity of the RBM (Figure 3C and 4). The few changes within the group of essential residues are almost exclusively hydrophobic "substitutions" (I428→L, L443→F, F460→Y, L472→F, Y484→Q), which should not affect either protein folding or the hACE2-interaction. At the same time, majority of the amino acid residues that are non-essential have "mutated" (Figure 4, RBM residues not labeled with short red lines). Judging from this sequence analysis alone, we were convinced early on that not only would the SARS-CoV-2 Spike protein bind hACE2 but also the binding would resemble, precisely, that between the original SARS Spike protein and hACE2[23]. Recent structural work has confirmed our prediction[39].

As elaborated below, the way that SARS-CoV-2 RBM resembles SARS-CoV RBM and the overall sequence conservation pattern between SARS-CoV-2 and ZC45/ZXC21 are highly unusual. Collectively, this suggests that portions of the SARS-CoV-2 genome have not been derived from natural quasi-species viral particle evolution.

If SARS-CoV-2 does indeed come from natural evolution, its RBM could have only been acquired in one of the two possible routes: 1) an ancient recombination event followed by convergent evolution or 2) a natural recombination event that occurred fairly recently.

In the first scenario, the ancestor of SARS-CoV-2, a ZC45/ZXC21-like bat coronavirus would have recombined and "swapped" its RBM with a coronavirus carrying a relatively "complete" RBM (in reference to SARS). This recombination would result in a novel ZC45/ZXC21-like coronavirus with all the gaps in its RBM "filled" (Figure 4). Subsequently, the virus would have to adapt extensively in its new host, where the ACE2 protein is highly homologous to hACE2. Random mutations across the genome would have to have occurred to eventually shape the RBM to its current form – resembling SARS-CoV RBM in a highly intelligent manner. However, this convergent evolution process would also result in the accumulation of a large amount of mutations in other parts of the genome, rendering the overall sequence identity relatively low. The high sequence identity between SARS-CoV-2 and ZC45/ZXC21 on various proteins (94-100% identity) do not support this scenario and, therefore, clearly indicates that SARS-CoV-2 carrying such an RBM cannot come from a ZC45/ZXC21-like bat coronavirus through this convergent evolutionary route.

In the second scenario, the ZC45/ZXC21-like coronavirus would have to have recently recombined and swapped its RBM with another coronavirus that had successfully adapted to bind an animal ACE2

highly homologous to hACE2. The likelihood of such an event depends, in part, on the general requirements of natural recombination: 1) that the two different viruses share significant sequence similarity; 2) that they must co-infect and be present in the same cell of the same animal; 3) that the recombinant virus would not be cleared by the host or make the host extinct; 4) that the recombinant virus eventually would have to become stable and transmissible within the host species.

In regard to this recent recombination scenario, the animal reservoir could not be bats because the ACE2 proteins in bats are not homologous enough to hACE2 and therefore the adaption would not be able to yield an RBM sequence as seen in SARS-CoV-2. This animal reservoir also could not be humans as the ZC45/ZXC21-like coronavirus would not be able to infect humans. In addition, there has been no evidence of any SARS-CoV-2 or SARS-CoV-2-like virus circulating in the human population prior to late 2019. Intriguingly, according to a recent bioinformatics study, SARS-CoV-2 was well-adapted for humans since the start of the outbreak[1].

Only one other possibility of natural evolution remains, which is that the ZC45/ZXC21-like virus and a coronavirus containing a SARS-like RBM could have recombined in an intermediate host where the ACE2 protein is homologous to hACE2. Several laboratories have reported that some of the Sunda pangolins smuggled into China from Malaysia carried coronaviruses, the receptor-binding domain (RBD) of which is almost identical to that of SARS-CoV-2[27-29,31]. They then went on to suggest that pangolins are the likely intermediate host for SARS-CoV-2[27-29,31]. However, recent independent reports have found significant flaws in this data[40-42]. Furthermore, contrary to these reports[27-29,31], no coronaviruses have been detected in Sunda pangolin samples collected for over a decade in Malaysia and Sabah between 2009 and 2019[43]. A recent study also showed that the RBD, which is shared between SARS-CoV-2 and the reported pangolin coronaviruses, binds to hACE2 ten times stronger than to the pangolin ACE2[2], further dismissing pangolins as the possible intermediate host. Finally, an *in silico* study, while echoing the notion that pangolins are not likely an intermediate host, also indicated that none of the animal ACE2 proteins examined in their study exhibited more favorable binding potential to the SARS-CoV-2 Spike protein than hACE2 did[3]. This last study virtually exempted all animals from their suspected roles as an intermediate host[3], which is consistent with the observation that SARS-CoV-2 was well-adapted for humans from the start of the outbreak[1]. This is significant because these findings collectively suggest that no intermediate host seems to exist for SARS-CoV-2, which at the very least diminishes the possibility of a recombinant event occurring in an intermediate host.

Even if we ignore the above evidence that no proper host exists for the recombination to take place and instead assume that such a host does exist, it is still highly unlikely that such a recombination event could occur in nature.

As we have described above, if natural recombination event is responsible for the appearance of SARS-CoV-2, then the ZC45/ZXC21-like virus and a coronavirus containing a SARS-like RBM would have to recombine in the same cell by swapping the S1/RBM, which is a rare form of recombination. Furthermore, since SARS has occurred only once in human history, it would be at least equally rare for nature to produce a virus that resembles SARS in such an intelligent manner – having an RBM that differs from the SARS RBM only at a few non-essential sites (Figure 4). The possibility that this unique SARS-like coronavirus would reside in the same cell with the ZC45/ZXC21-like ancestor virus and the two viruses would recombine in the "RBM-swapping" fashion is extremely low. Importantly, this, and the other recombination event described below in section 1.3 (even more impossible to occur in nature), would both have to happen to produce a Spike as seen in SARS-CoV-2.

While the above evidence and analyses together appear to disapprove a natural origin of SARS-CoV-2's RBM, abundant literature shows that gain-of-function research, where the Spike protein of a coronavirus was specifically engineered, has repeatedly led to the successful generation of human-infecting coronaviruses from coronaviruses of non-human origin[44-47].

Record also shows that research laboratories, for example, the Wuhan Institute of Virology (WIV), have successfully carried out such studies working with US researchers[45] and also working alone[47]. In addition, the WIV has engaged in decades-long coronavirus surveillance studies and therefore owns the world's largest collection of coronaviruses. Evidently, the technical barrier is non-existent for the WIV and other related laboratories to carry out and succeed in such Spike/RBM engineering and gain-of-function research.

**A SARS-CoV-2**

```
                                                         EcoRI
                                                       W   N   S
tataattata aattaccaga tgattttaca ggctgcgtta tagcttggaa ttctaacaat   1320
cttgattcta aggttggtgg taattataat tacctgtata gattgtttag gaagtctaat   1380
ctcaaacctt ttgagagaga tatttcaact gaaatctatc aggccggtag cacaccttgt   1440
aatggtgttg aaggttttaa ttgttacttt cctttacaat catatggttt ccaacccact   1500
aatggtgttg gttaccaacc atacagagta gtagtacttt cttttgaact tctacatgca   1560
              G  Y  Q
              BstEII
```

**B ZC45**

```
                                      W   N   T
ttacctgatg attttacagg ttgtgtcata gcttggaaca ctgccaaaca ggatgtaggt   1320
aattatttct acaggtctca tcgttctacc aaattgaaac catttgaaag agatctttcc   1380
tcagacgaga atggtgtccg tacacttagt acttatgact tcaaccctaa tgtaccactt   1440
gaataccaag ctacaagggt tgttgttttg tcatttgagc ttctaaatgc accagctaca   1500
 E  Y  Q
```

**C**

```
SARS-CoV-2_RBM_EcoRI/BstEII    KIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG
Shang, J. et al. 2020          KIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG
Ren, W. et al. 2008            VIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPD
                                ************ ***:***:.*:*:.  ***** ** :*:.:*.*******.  :...

SARS-CoV-2_RBM_EcoRI/BstEII    STPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKN
Shang, J. et al. 2020          STPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKN
Ren, W. et al. 2008            GKPCTP-PALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKN
                               ..**.    .:***:**:.*** .*.*:***************:********* **:*:**
```

***Figure 5. Two restriction sites are present at either end of the RBM of SARS-CoV-2, providing convenience for replacing the RBM within the spike gene.*** *A. Nucleotide sequence of the RBM of SARS-CoV-2 (Wuhan-Hu-1). An EcoRI site is found at the 5'-end of the RBM and a BstEII site at the 3'-end. B. Although these two restriction sites do not exist in the original spike gene of ZC45, they can be conveniently introduced given that the sequence discrepancy is small (2 nucleotides) in either case. C. Amino acid sequence alignment with the RBM region highlighted (color and underscore). The RBM highlighted in orange (top) is what is defined by the EcoRI and BstEII sites in the SARS-CoV-2 (Wuhan-Hu-1) spike. The RBM highlighted in magenta (middle) is the region swapped by Dr. Fang Li and colleagues into a SARS Spike backbone[39]. The RBM highlighted in blue (bottom) is from the Spike protein (RBM: 424-494) of SARS-BJ01 (AY278488.2), which was swapped by the Shi lab into the Spike proteins of different bat coronaviruses replacing the corresponding segments[47].*

Strikingly, consistent with the RBM engineering theory, we have identified two unique restriction sites, EcoRI and BstEII, at either end of the *RBM* of the SARS-CoV-2 genome, respectively (Figure 5A). These two sites, which are popular choices of everyday molecular cloning, do not exist in the rest of this *spike* gene. This particular setting makes it extremely convenient to swap the *RBM* within *spike*, providing a quick way to test different RBMs and the corresponding Spike proteins.

Such EcoRI and BstEII sites do not exist in the *spike* genes of other β coronaviruses, which strongly indicates that they were unnatural and were specifically introduced into this *spike* gene of SARS-CoV-2 for the convenience of manipulating the critical RBM. Although ZC45 *spike* also does not have these two sites (Figure 5B), they can be introduced very easily as described in part 2 of this report.

It is noteworthy that introduction of the EcoRI site here would change the corresponding amino acids from -*WNT*- to -*WNS*- (Figure 5AB). As far as we know, all SARS and SARS-like bat coronaviruses exclusively carry a *T* (threonine) residue at this location. SARS-CoV-2 is the only exception in that this *T* has mutated to an *S* (serine), save the suspicious RaTG13 and pangolin coronaviruses published after the outbreak[48].

Once the restriction sites were successfully introduced, the *RBM* segment could be swapped conveniently using routine restriction enzyme digestion and ligation. Although alternative cloning techniques may leave no trace of genetic manipulation (Gibson assembly as one example), this old-fashioned approach could be chosen because it offers a great level of convenience in swapping this critical *RBM*.

Given that RBM fully dictates hACE2-binding and that the SARS RBM-hACE2 binding was fully characterized by high-resolution structures (Figure 3)[37,38], this RBM-only swap would not be any riskier than the full Spike swap. In fact, the feasibility of this RBM-swap strategy has been proven[39,47]. In 2008, Dr. Zhengli Shi's group swapped a SARS RBM into the Spike proteins of several SARS-like bat coronaviruses after introducing a restriction site into a codon-optimized *spike* gene (Figure 5C)[47]. They then validated the binding of the resulted chimeric Spike proteins with hACE2. Furthermore, in a recent publication, the RBM of SARS-CoV-2 was swapped into the receptor-binding domain (RBD) of SARS-CoV, resulting in a chimeric RBD fully functional in binding hACE2 (Figure 5C)[39]. Strikingly, in both cases, the manipulated RBM segments resemble almost exactly the RBM defined by the positions of the EcoRI and BstEII sites (Figure 5C). Although cloning details are lacking in both publications[39,47], it is conceivable that the actual restriction sites may vary depending on the *spike* gene receiving the RBM insertion as well as the convenience in introducing unique restriction site(s) in regions of interest. It is noteworthy that the corresponding author of this recent publication[39], Dr. Fang Li, has been an active collaborator of Dr. Zhengli Shi since 2010[49-53]. Dr. Li was the first person in the world to have structurally elucidated the binding between SARS-CoV RBD and hACE2[38] and has been the leading expert in the structural understanding of Spike-ACE2 interactions[38,39,53-56]. The striking finding of EcoRI and BstEII restriction sites at either end of the SARS-CoV-2 RBM, respectively, and the fact that the same RBM region has been swapped both by Dr. Shi and by her long-term collaborator, respectively, using restriction enzyme digestion methods are unlikely a coincidence. **Rather, it is the smoking gun proving that the RBM/Spike of SARS-CoV-2 is a product of genetic manipulation.**

Although it may be convenient to copy the exact sequence of SARS RBM, it would be too clear a sign of artificial design and manipulation. The more deceiving approach would be to change a few non-essential residues, while preserving the ones critical for binding. This design could be well-guided by the high-resolution structures (Figure 3)[37,38]. This way, when the overall sequence of the RBM would appear

to be more distinct from that of the SARS RBM, the hACE2-binding ability would be well-preserved. We believe that all of the crucial residues (residues labeled with red sticks in Figure 4, which are the same residues shown in sticks in Figure 3C) should have been "kept". As described earlier, while some should be direct preservation, some should have been switched to residues with similar properties, which would not disrupt hACE2-binding and may even strengthen the association further. Importantly, changes might have been made intentionally at non-essential sites, making it less like a "copy and paste" of the SARS RBM.

### 1.3 An unusual furin-cleavage site is present in the Spike protein of SARS-CoV-2 and is associated with the augmented virulence of the virus

Another unique motif in the Spike protein of SARS-CoV-2 is a polybasic furin-cleavage site located at the S1/S2 junction (Figure 4, segment in between two green lines). Such a site can be recognized and cleaved by the furin protease. Within the lineage B of β coronaviruses and with the exception of SARS-CoV-2, no viruses contain a furin-cleavage site at the S1/S2 junction (Figure 6)[57]. In contrast, furin-cleavage site at this location has been observed in other groups of coronaviruses[57,58]. Certain selective pressure seems to be in place that prevents the lineage B of β coronaviruses from acquiring or maintaining such a site in nature.

```
Human SARS-CoV BJ01                655 - GICASYHTVSLL----RSTS - 670
Human SARS-CoV CUHK-W1             655 - GICASYHTVSLL----RSTS - 670
Human SARS-CoV Tor2                655 - GICASYHTVSLL----RSTS - 670
Human SARS-CoV Frankfurt-1         655 - GICASYHTVSLL----RSTS - 670
Human SARS-CoV Urbani              655 - GICASYHTVSLL----RSTS - 670
Civet SARS-CoV civet020            655 - GICASYHTVSSL----RSTS - 670
Civet SARS-CoV sz16                655 - GICASYHTVSSL----RSTS - 670
Racoon dog SARS-CoV A030           655 - GICASYHTVSSL----RSTS - 670
SARS-CoV-2                         669 - GICASYQTQTNSPRRARSVA - 688
Pangolin CoV MP789                 n/a - GICASYQTQTNS----RSVS - n/a
Bat SARSr-CoV RaTG13               669 - GICASYQTQTNS----RSVA - 684
Bat SARSr-CoV LYRa11               659 - GICASYHTASLL----RNTD - 674
Bat SARSr-CoV LYRa3                659 - GICASYHTASLL----RNTG - 674
Bat SARSr-CoV RsSHC014             656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs4084               656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV WIV1                 656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs3367               656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs7327               656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs9401               656 - GICASYHTVSSL----RSTS - 671
Bat SARSr-CoV Rs4231               655 - GICASYHTVSSL----RSTS - 670
Bat SARSr-CoV WIV16                655 - GICASYHTVSSL----RSTS - 670
Bat SARSr-CoV Rs4874               655 - GICASYHTVSSL----RSTS - 670
Bat SARSr-CoV ZC45                 646 - GICASYHTASIL----RSTS - 661
Bat SARSr-CoV ZXC21                645 - GICASYHTASIL----RSTG - 660
Bat SARSr-CoV Rf4092               634 - GICASYHTASTL----RGVG - 649
Bat SARSr-CoV Rf/JL2012            636 - GICASYHTASLL----RSTG - 651
Bat SARSr-CoV JTMC15               636 - GICASYHTASLL----RSTG - 651
Bat SARSr-CoV 16BO133              636 - GICASYHTASLL----RSTG - 651
Bat SARSr-CoV B15-21               636 - GICASYHTASLL----RSTG - 651
Bat SARSr-CoV YN2013               633 - GICASYHTASTL----RSIG - 648
Bat SARSr-CoV Anlong-103           633 - GICASYHTASTL----RSVG - 648
Bat SARSr-CoV Rp/Shaanxi2011       640 - GICASYHTASVL----RSTG - 655
Bat SARSr-CoV Rs/HuB2013           641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV YNLF/34C             641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV YNLF/31C             641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV Rf1                  641 - GICASYHTASHL----RSTG - 656
Bat SARSr-CoV 273                  641 - GICASYHTASHL----RSTG - 656
Bat SARSr-CoV Rf/SX2013            639 - GICASYHTASLL----RSTG - 654
Bat SARSr-CoV Rf/HeB2013           641 - GICASYHTASLL----RSTG - 656
Bat SARSr-CoV Cp/Yunnan2011        641 - GICASYHTASLL----RNTG - 656
Bat SARSr-CoV Rs672                641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV Rs4255               641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV 4081                 641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV Rm1                  641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV 279                  641 - GICASYHTASVL----RSTG - 656
Bat SARSr-CoV Rs/GX2013            642 - GICASYHTASVL----RSTG - 657
Bat SARSr-CoV Rs806                641 - GICASYHTASLL----RSTG - 656
Bat SARSr-CoV HKU3-1               642 - GICASYHTASVL----RSTG - 657
Bat SARSr-CoV Longquan-140         642 - GICASYHTASVL----RSTG - 657
Bat SARSr-CoV Rp3                  641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV Rs4247               642 - GICASYHTASTL----RSVG - 657
Bat SARSr-CoV Rs4237               641 - GICASYHTASTL----RSVG - 656
Bat SARSr-CoV As6526               641 - GICASYHTASTL----RSVG - 656
```

***Figure 6. Furin-cleavage site found at the S1/S2 junction of Spike is unique to SARS-CoV-2 and absent in other lineage B β coronaviruses.*** *Figure reproduced from Hoffmann, et al*[57].

As previously described, during the cell entry process, the Spike protein is first cleaved at the S1/S2 junction. This step, and a subsequent cleavage downstream that exposes the fusion peptide, are both mediated by host proteases. The presence or absence of these proteases in different cell types greatly affects the cell tropism and presumably the pathogenicity of the viral infection. Unlike other proteases, furin protease is widely expressed in many types of cells and is present at multiple cellular and extracellular locations. Importantly, the introduction of a furin-cleavage site at the S1/S2 junction could significantly enhance the infectivity of a virus as well as greatly expand its cell tropism — a phenomenon well-documented in both influenza viruses and other coronaviruses[59-65].

If we leave aside the fact that no furin-cleavage site is found in any lineage B β coronavirus in nature and instead assume that this site in SARS-CoV-2 is a result of natural evolution, then only one evolutionary pathway is possible, which is that the furin-cleavage site has to be derived from a homologous recombination event. Specifically, an ancestor β coronavirus containing no furin-cleavage site would have to recombine with a closely related coronavirus that does contain a furin-cleavage site.

However, two facts disfavor this possibility. First, although some coronaviruses from other groups or lineages do contain polybasic furin-cleavage sites, none of them contains the exact polybasic sequence present in SARS-CoV-2 (*-PRRAR/SVA-*). Second, between SARS-CoV-2 and any coronavirus containing a legitimate furin-cleavage site, the sequence identity on Spike is no more than 40%[66]. Such a low level of sequence identity rules out the possibility of a successful homologous recombination ever occurring between the ancestors of these viruses. Therefore, the furin-cleavage site within the SARS-CoV-2 Spike protein is unlikely to be of natural origin and instead should be a result of laboratory modification.

Consistent with this claim, a close examination of the nucleotide sequence of the furin-cleavage site in SARS-CoV-2 *spike* has revealed that the two consecutive Arg residues within the inserted sequence (-PRRA-) are both coded by the rare codon CGG (least used codon for Arg in SARS-CoV-2) (Figure 7)[8]. In fact, this *CGGCGG* arrangement is the only instance found in the SARS-CoV-2 genome where this rare codon is used in tandem. This observation strongly suggests that this furin-cleavage site should be a result of genetic engineering. Adding to the suspicion, a *FauI* restriction site is formulated by the codon choices here, suggesting the possibility that the *restriction fragment length polymorphism,* a technique that a WIV lab is proficient at[67], could have been involved. There, the fragmentation pattern resulted from *FauI* digestion could be used to monitor the preservation of the furin-cleavage site in Spike as this furin-cleavage site is prone to deletions *in vitro*[68,69]. Specifically, RT-PCR on the *spike* gene of the recovered viruses from cell cultures or laboratory animals could be carried out, the product of which would be subjected to *FauI* digestion. Viruses retaining or losing the furin-cleavage site would then yield distinct patterns, allowing convenient tracking of the virus(es) of interest.

**FauI**

```
tat cag act cag act aat tct cct cgg cgg gca cgt agt gta gct agt caa tcc atc att
 Y   Q   T   Q   T   N   S   P   R   R   A   R   S   V   A   S   Q   S   I   I
```

***Figure 7. Two consecutive Arg residues in the -PRRA- insertion at the S1/S2 junction of SARS-CoV-2 Spike are both coded by a rare codon, CGG.*** *A FauI restriction site, 5'-$(N)_6$GCGGG-3', is embedded in the coding sequence of the "inserted" PRRA segment, which may be used as a marker to monitor the preservation of the introduced furin-cleavage site.*

In addition, although no known coronaviruses contain the exact sequence of *-PRRAR/SVA-* that is present in the SARS-CoV-2 Spike protein, a similar *-RRAR/AR-* sequence has been observed at the S1/S2 junction of the Spike protein in a rodent coronavirus, AcCoV-JC34, which was published by Dr. Zhengli

Shi in 2017[70]. It is evident that the legitimacy of *-RRAR-* as a functional furin-cleavage site has been known to the WIV experts since 2017.

The evidence collectively suggests that the furin-cleavage site in the SARS-CoV-2 Spike protein may not have come from nature and could be the result of genetic manipulation. The purpose of this manipulation could have been to assess any potential enhancement of the infectivity and pathogenicity of the laboratory-made coronavirus[59-64]. Indeed, recent studies have confirmed that the furin-cleavage site does confer significant pathogenic advantages to SARS-CoV-2[57,68].

### 1.4 Summary

Evidence presented in this part reveals that certain aspects of the SARS-CoV-2 genome are extremely difficult to reconcile to being a result of natural evolution. The alternative theory we suggest is that the virus may have been created by using ZC45/ZXC21 bat coronavirus(es) as the backbone and/or template. The Spike protein, especially the RBM within it, should have been artificially manipulated, upon which the virus has acquired the ability to bind hACE2 and infect humans. This is supported by the finding of a unique restriction enzyme digestion site at either end of the RBM. An unusual furin-cleavage site may have been introduced and inserted at the S1/S2 junction of the Spike protein, which contributes to the increased virulence and pathogenicity of the virus. These transformations have then staged the SARS-CoV-2 virus to eventually become a highly-transmissible, onset-hidden, lethal, sequelae-unclear, and massively disruptive pathogen.

Evidently, the possibility that SARS-CoV-2 could have been created through gain-of-function manipulations at the WIV is significant and should be investigated thoroughly and independently.

## 2. Delineation of a synthetic route of SARS-CoV-2

In the second part of this report, we describe a synthetic route of creating SARS-CoV-2 in a laboratory setting. It is postulated based on substantial literature support as well as genetic evidence present in the SARS-CoV-2 genome. Although steps presented herein should not be viewed as exactly those taken, we believe that key processes should not be much different. Importantly, our work here should serve as a demonstration of how SARS-CoV-2 can be designed and created conveniently in research laboratories by following proven concepts and using well-established techniques.

Importantly, research labs, both in Hong Kong and in mainland China, are leading the world in coronavirus research, both in terms of resources and on the research outputs. The latter is evidenced not only by the large number of publications that they have produced over the past two decades but also by their milestone achievements in the field: they were the first to identify civets as the intermediate host for SARS-CoV and isolated the first strain of the virus[71]; they were the first to uncover that SARS-CoV originated from bats[72,73]; they revealed for the first time the antibody-dependent enhancement (ADE) of SARS-CoV infections[74]; they have contributed significantly in understanding MERS in all domains (zoonosis, virology, and clinical studies)[75-79]; they made several breakthroughs in SARS-CoV-2 research[18,35,80]. Last but not least, they have the world's largest collection of coronaviruses (genomic sequences and live viruses). The knowledge, expertise, and resources are all readily available within the Hong Kong and mainland research laboratories (they collaborate extensively) to carry out and accomplish the work described below.

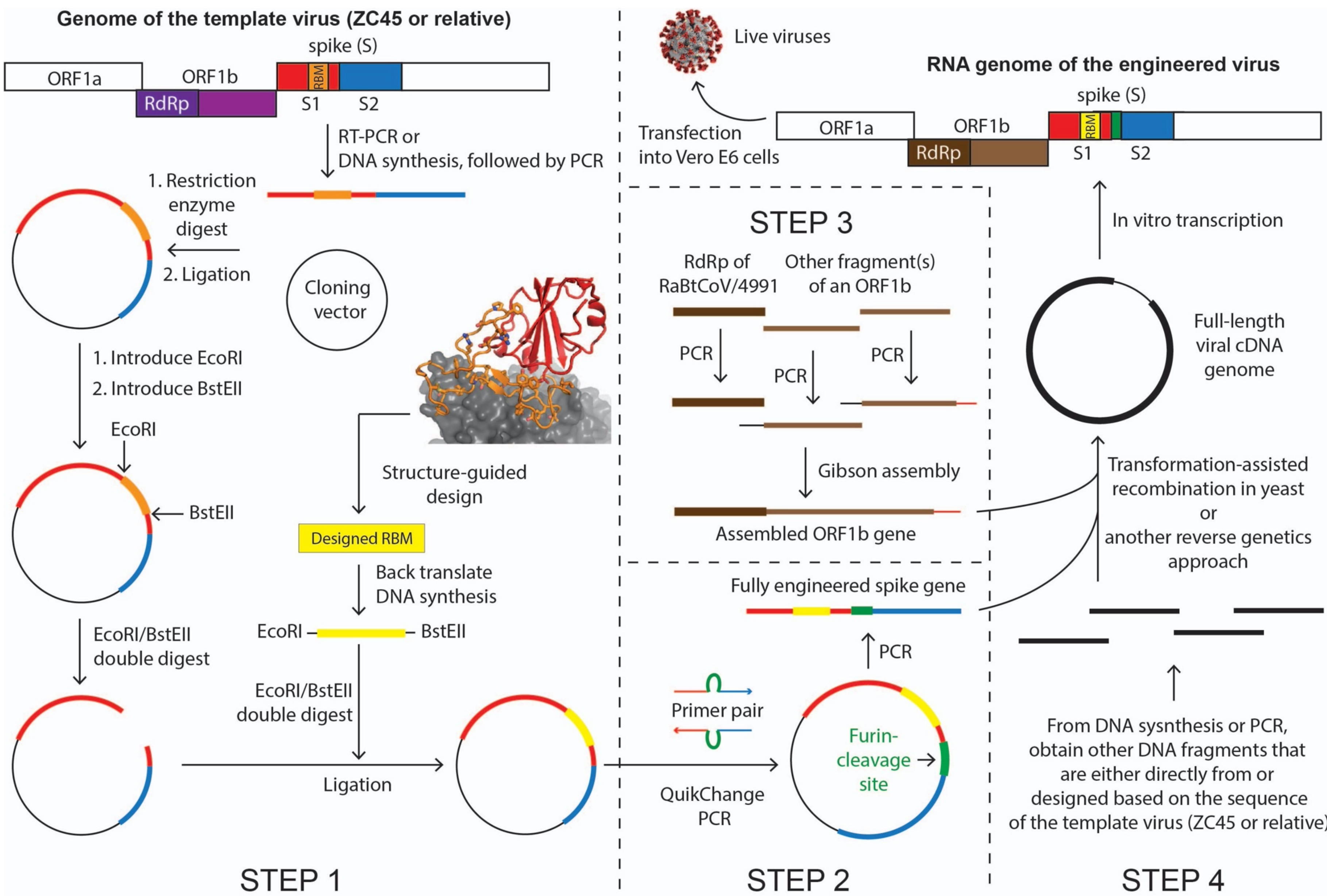



***Figure 8. Diagram of a possible synthetic route of the laboratory-creation of SARS-CoV-2.***

## 2.1 Possible scheme in designing the laboratory-creation of the novel coronavirus

In this sub-section, we outline the possible overall strategy and major considerations that may have been formulated at the designing stage of the project.

To engineer and create a human-targeting coronavirus, they would have to pick a bat coronavirus as the template/backbone. This can be conveniently done because many research labs have been actively collecting bat coronaviruses over the past two decades[32,33,70,72,81-85]. However, this template virus ideally should not be one from Dr. Zhengli Shi's collections, considering that she is widely known to have been engaged in gain-of-function studies on coronaviruses. Therefore, ZC45 and/or ZXC21, novel bat coronaviruses discovered and owned by military laboratories[33], would be suitable as the template/backbone. It is also possible that these military laboratories had discovered other closely related viruses from the same location and kept some unpublished. Therefore, the actual template could be ZC45, or ZXC21, or a close relative of them. The postulated pathway described below would be the same regardless of which one of the three was the actual template.

Once they have chosen a template virus, they would first need to engineer, through molecular cloning, the Spike protein so that it can bind hACE2. The concept and cloning techniques involved in this manipulation have been well-documented in the literature[44-46,84,86]. With almost no risk of failing, the template bat virus could then be converted to a coronavirus that can bind hACE2 and infect humans[44-46].

Second, they would use molecular cloning to introduce a furin-cleavage site at the S1/S2 junction of Spike. This manipulation, based on known knowledge[60,61,65], would likely produce a strain of coronavirus that is a more infectious and pathogenic.

Third, they would produce an *ORF1b* gene construct. The *ORF1b* gene encodes the polyprotein Orf1b, which is processed post-translationally to produce individual viral proteins: RNA-dependent RNA polymerase (RdRp), helicase, guanidine-N7 methyltransferase, uridylate-specific endoribonuclease, and 2'-O-methyltransferase. All of these proteins are parts of the replication machinery of the virus. Among them, the RdRp protein is the most crucial one and is highly conserved among coronaviruses. Importantly, Dr. Zhengli Shi's laboratory uses a PCR protocol, which amplifies a particular fragment of the *RdRp* gene, as their primary method to detect the presence of coronaviruses in raw samples (bat fecal swap, feces, etc). As a result of this practice, the Shi group has documented the sequence information of this short segment of *RdRp* for all coronaviruses that they have successfully detected and/or collected.

Here, the genetic manipulation is less demanding or complicated because Orf1b is conserved and likely Orf1b from any β coronavirus would be competent enough to do the work. However, we believe that they would want to introduce a particular Orf1b into the virus for one of the two possible reasons:

1. Since many phylogenetic analyses categorize coronaviruses based on the sequence similarity of the *RdRp* gene only[18,31,35,83,87], having a different *RdRp* in the genome therefore could ensure that SARS-CoV-2 and ZC45/ZXC21 are separated into different groups/sub-lineages in phylogenetic studies. Choosing an *RdRp* gene, however, is convenient because the short *RdRp* segment sequence has been recorded for all coronaviruses ever collected/detected. Their final choice was the *RdRp* sequence from bat coronavirus RaBtCoV/4991, which was discovered in 2013. For RaBtCoV/4991, the only information ever published was the sequence of its short *RdRp* segment[83], while neither its full genomic sequence nor virus isolation were ever reported. After amplifying the *RdRp* segment (or the whole *ORF1b* gene) of RaBatCoV/4991, they would have then used it for subsequent assembly and creation of the genome of SARS-CoV-2. Small changes in the *RdRp*

sequence could either be introduced at the beginning (through DNA synthesis) or be generated *via* passages later on. On a separate track, when they were engaged in the fabrication of the RaTG13 sequence, they could have started with the short *RdRp* segment of RaBtCoV/4991 without introducing any changes to its sequence, resulting in a 100% nucleotide sequence identity between the two viruses on this short *RdRp* segment[83]. This RaTG13 virus could then be claimed to have been discovered back in 2013.

2. The RdRp protein from RaBatCoV/4991 is unique in that it is superior than RdRp from any other β coronavirus for developing antiviral drugs. RdRp has no homologs in human cells, which makes this essential viral enzyme a highly desirable target for antiviral development. As an example, *Remedesivir*, which is currently undergoing clinical trials, targets RdRp. When creating a novel and human-targeting virus, they would be interested in developing the antidote as well. Even though drug discovery like this may not be easily achieved, it is reasonable for them to intentionally incorporate a RdRp that is more amenable for antiviral drug development.

Fourth, they would use reverse genetics to assemble the gene fragments of *spike*, *ORF1b*, and the rest of the template ZC45 into a cDNA version of the viral genome. They would then carry out *in vitro* transcription to obtain the viral RNA genome. Transfection of the RNA genome into cells would allow the recovery of live and infectious viruses with the desired artificial genome.

Fifth, they would carry out characterization and optimization of the virus strain(s) to improve the fitness, infectivity, and overall adaptation using serial passage *in vivo*. One or several viral strains that meet certain criteria would then be obtained as the final product(s).

### 2.2 A postulated synthetic route for the creation of SARS-CoV-2

In this sub-section, we describe in more details how each step could be carried out in a laboratory setting using available materials and routine molecular, cellular, and virologic techniques. A diagram of this process is shown in Figure 8. We estimate that the whole process could be completed in approximately 6 months.

Step 1: Engineering the RBM of the Spike for hACE2-binding (1.5 months)

The Spike protein of a bat coronavirus is either incapable of or inefficient in binding hACE2 due to the missing of important residues within its RBM. This can be exemplified by the RBM of the template virus ZC45 (Figure 4). The first and most critical step in the creation of SARS-CoV-2 is to engineer the Spike so that it acquires the ability to bind hACE2. As evidenced in the literature, such manipulations have been carried out repeatedly in research laboratories since 2008[44], which successfully yielded engineered coronaviruses with the ability to infect human cells[44-46,88,89]. Although there are many possible ways that one can engineer the Spike protein, we believe that what was actually undertaken was that they replaced the original RBM with a designed and possibly optimized RBM using SARS' RBM as a guide. As described in part 1, this theory is supported by our observation that two unique restriction sites, EcoRI and BstEII, exist at either end of the *RBM* in the SARS-CoV-2 genome (Figure 5A) and by the fact that such RBM-swap has been successfully carried out by Dr. Zhengli Shi and by her long-term collaborator and structure biology expert, Dr. Fang Li[39,47].

Although ZC45 *spike* does not contain these two restriction sites (Figure 5B), they can be introduced very easily. The original *spike* gene would be either amplified with RT-PCR or obtained through DNA synthesis (some changes could be safely introduced to certain variable regions of the sequence) followed by PCR. The gene would then be cloned into a plasmid using restriction sites other than EcoRI and BstEII.

Once in the plasmid, the *spike* gene can be modified easily. First, an EcoRI site can be introduced by converting the highlighted "gaacac" sequence (Figure 5B) to the desired "gaattc" (Figure 5A). The difference between them are two consecutive nucleotides. Using the commercially available QuikChange Site-Directed Mutagenesis kit, such a di-nucleotide mutation can be generated in no more than one week. Subsequently, the BstEII site could be similarly introduced at the other end of the *RBM*. Specifically, the "gaatacc" sequence (Figure 5B) would be converted to the desired "ggttacc" (Figure 5A), which would similarly require a week of time.

Once these restriction sites, which are unique within the *spike* gene of SARS-CoV-2, were successfully introduced, different *RBM* segments could be swapped in conveniently and the resulting Spike protein subsequently evaluated using established assays.

As described in part 1, the design of an RBM segment could be well-guided by the high-resolution structures (Figure 3)[37,38], yielding a sequence that resembles the SARS RBM in an intelligent manner. When carrying out the structure-guided design of the RBM, they would have followed the routine and generated a few (for example a dozen) such RBMs with the hope that some specific variant(s) may be superior than others in binding hACE2. Once the design was finished, they could have each of the designed *RBM* genes commercially synthesized (quick and very affordable) with an EcoRI site at the 5'-end and a BstEII site at the 3'-end. These novel *RBM* genes could then be cloned into the *spike* gene, respectively. The gene synthesis and subsequent cloning, which could be done in a batch mode for the small library of designed RBMs, would take approximately one month.

These engineered Spike proteins might then be tested for hACE2-binding using the established pseudotype virus infection assays[45,49,50]. The engineered Spike with good to exceptional binding affinities would be selected. (Although not necessary, directed evolution could be involved here (error-prone PCR on the *RBM* gene), coupled with either an *in vitro* binding assay[39,90] or a pseudotype virus infection assay[45,49,50], to obtain an RBM that binds hACE2 with exceptional affinity.)

Given the abundance of literature on Spike engineering[44-46,84,86] and the available high-resolution structures of the Spike-hACE2 complex[37,38], the success of this step would be very much guaranteed. By the end of this step, as desired, a novel *spike* gene would be obtained, which encodes a novel Spike protein capable of binding hACE2 with high affinity.

<u>Step 2: Engineering a furin-cleavage site at the S1/S2 junction (0.5 month)</u>

The product from Step 1, a plasmid containing the engineered *spike*, would be further modified to include a furin-cleavage site (segment indicated by green lines in Figure 4) at the S1/S2 junction. This short stretch of gene sequence can be conveniently inserted using several routine cloning techniques, including QuikChange Site-Directed PCR[60], overlap PCR followed by restriction enzyme digestion and ligation[91], or Gibson assembly. None of these techniques would leave any trace in the sequence. Whichever cloning method was the choice, the inserted gene piece would be included in the primers, which would be designed, synthesized, and used in the cloning. This step, leading to a further modified Spike with the furin-cleavage site added at the S1/S2 junction, could be completed in no more than two weeks.

<u>Step 3: Obtain an *ORF1b* gene that contains the sequence of the short *RdRp* segment from RaBtCoV/4991 (1 month, yet can be carried out concurrently with Steps 1 and 2)</u>

Unlike the engineering of Spike, no complicated design is needed here, except that the *RdRp* gene segment from RaBtCoV/4991 would need to be included. Gibson assembly could have been used here. In this technique, several fragments, each adjacent pair sharing 20-40 bp overlap, are combined together in one simple reaction to assemble a long DNA product. Two or three fragments, each covering a significant section of the *ORF1b* gene, would be selected based on known bat coronavirus sequences. One of these fragments would be the *RdRp* segment of RaBtCoV/4991[83]. Each fragment would be PCR amplified with proper overlap regions introduced in the primers. Finally, all purified fragments would be pooled in equimolar concentrations and added to the Gibson reaction mixture, which, after a short incubation, would yield the desired *ORF1b* gene in whole.

Step 4: Produce the designed viral genome using reverse genetics and recover live viruses (0.5 month)

Reverse genetics have been frequently used in assembling whole viral genomes, including coronavirus genomes[67,92-96]. The most recent example is the reconstruction of the SARS-CoV-2 genome using the *transformation-assisted recombination in yeast*[97]. Using this method, the Swiss group assembled the entire viral genome and produced live viruses in just one week[97]. This efficient technique, which would not leave any trace of artificial manipulation in the created viral genome, has been available since 2017[98,99]. In addition to the engineered *spike* gene (from steps 1 and 2) and the *ORF1b* gene (from step 3), other fragments covering the rest of the genome would be obtained either through RT-PCR amplification from the template virus or through DNA synthesis by following a sequence slightly altered from that of the template virus. We believe that the latter approach was more likely as it would allow sequence changes introduced into the variable regions of less conserved proteins, the process of which could be easily guided by multiple sequence alignments. The amino acid sequences of more conserved functions, such as that of the E protein, might have been left unchanged. All DNA fragments would then be pooled together and transformed into yeast, where the cDNA version of the SARS-CoV-2 genome would be assembled *via* transformation-assisted recombination. Of course, an alternative method of reverse genetics, one of which the WIV has successfully used in the past[67], could also be employed[67,92-96,100]. Although some earlier reverse genetics approaches may leave restriction sites at where different fragments would be joined, these traces would be hard to detect as the exact site of ligation can be anywhere in the ~30kb genome. Either way, a cDNA version of the viral genome would be obtained from the reverse genetics experiment. Subsequently, *in vitro* transcription using the cDNA as the template would yield the viral RNA genome, which upon transfection into Vero E6 cells would allow the production of live viruses bearing all of the designed properties.

Step 5: Optimize the virus for fitness and improve its hACE2-binding affinity *in vivo* (2.5-3 months)

Virus recovered from step 4 needs to be further adapted undergoing the classic experiment – serial passage in laboratory animals[101]. This final step would validate the virus' fitness and ensure its receptor-oriented adaptation toward its intended host, which, according to the analyses above, should be human. Importantly, the RBM and the furin-cleavage site, which were introduced into the Spike protein separately, would now be optimized together as one functional unit. Among various available animal models (e.g. mice, hamsters, ferrets, and monkeys) for coronaviruses, hACE2 transgenic mice (hACE2-mice) should be the most proper and convenient choice here. This animal model has been established during the study of SARS-CoV and has been available in the Jackson Laboratory for many years[102-104].

The procedure of serial passage is straightforward. Briefly, the selected viral strain from step 4, a precursor of SARS-CoV-2, would be intranasally inoculated into a group of anaesthetized hACE2-mice. Around 2-3 days post infection, the virus in lungs would usually amplify to a peak titer. The mice would

then be sacrificed and the lungs homogenized. Usually, the mouse-lung supernatant, which carries the highest viral load, would be used to extract the candidate virus for the next round of passage. After approximately 10~15 rounds of passage, the hACE2-binding affinity, the infection efficiency, and the lethality of the viral strain would be sufficiently enhanced and the viral genome stabilized[101]. Finally, after a series of characterization experiments (e.g. viral kinetics assay, antibodies response assay, symptom observation and pathology examination), the final product, SARS-CoV-2, would be obtained, concluding the whole creation process. From this point on, this viral pathogen could be amplified (most probably using Vero E6 cells) and produced routinely.

It is noteworthy that, based on the work done on SARS-CoV, the hACE2-mice, although suitable for SARS-CoV-2 adaptation, is not a good model to reflect the virus' transmissibility and associated clinical symptoms in humans. We believe that those scientists might not have used a proper animal model (such as the golden Syrian hamster) for testing the transmissibility of SARS-CoV-2 before the outbreak of COVID-19. If they had done this experiment with a proper animal model, the highly contagious nature of SARS-CoV-2 would be extremely evident and consequently SARS-CoV-2 would not have been described as "not causing human-to-human transmission" at the start of the outbreak.

We also speculate that the extensive laboratory-adaptation, which is oriented toward enhanced transmissibility and lethality, may have driven the virus too far. As a result, SARS-CoV-2 might have lost the capacity to attenuate on both transmissibility and lethality during its current adaptation in the human population. This hypothesis is consistent with the lack of apparent attenuation of SARS-CoV-2 so far despite its great prevalence and with the observation that a recently emerged, predominant variant only shows improved transmissibility[105-108].

Serial passage is a quick and intensive process, where the adaptation of the virus is accelerated. Although intended to mimic natural evolution, serial passage is much more limited in both time and scale. As a result, less random mutations would be expected in serial passage than in natural evolution. This is particularly true for conserved viral proteins, such as the E protein. Critical in viral replication, the E protein is a determinant of virulence and engineering of it may render SARS-CoV-2 attenuated[109-111] Therefore, at the initial assembly stage, these scientists might have decided to keep the amino acid sequence of the E protein unchanged from that of ZC45/ZXC21. Due to the conserved nature of the E protein and the limitations of serial passage, no amino acid mutation actually occurred, resulting in a 100% sequence identity on the E protein between SARS-CoV-2 and ZC45/ZXC21. The same could have happened to the marks of molecular cloning (restriction sites flanking the RBM). Serial passage, which should have partially naturalized the SARS-CoV-2 genome, might not have removed all signs of artificial manipulation.

## 3. Final remarks

Many questions remain unanswered about the origin of SARS-CoV-2. Prominent virologists have implicated in a *Nature Medicine* letter that laboratory escape, while not being entirely ruled out, was unlikely and that no sign of genetic manipulation is present in the SARS-CoV-2 genome[4]. However, here we show that genetic evidence within the *spike* gene of SARS-CoV-2 genome (restriction sites flanking the *RBM*; tandem rare codons used at the inserted furin-cleavage site) does exist and suggests that the SARS-CoV-2 genome should be a product of genetic manipulation. Furthermore, the proven concepts, well-established techniques, and knowledge and expertise are all in place for the convenient creation of this novel coronavirus in a short period of time.

Motives aside, the following facts about SARS-CoV-2 are well-supported:

1. If it was a laboratory product, the most critical element in its creation, the backbone/template virus (ZC45/ZXC21), is owned by military research laboratories.
2. The genome sequence of SARS-CoV-2 has likely undergone genetic engineering, through which the virus has gained the ability to target humans with enhanced virulence and infectivity.
3. The characteristics and pathogenic effects of SARS-CoV-2 are unprecedented. The virus is highly transmissible, onset-hidden, multi-organ targeting, sequelae-unclear, lethal, and associated with various symptoms and complications.
4. SARS-CoV-2 caused a world-wide pandemic, taking hundreds of thousands of lives and shutting down the global economy. It has a destructive power like no other.

Judging from the evidence that we and others have gathered, we believe that finding the origin of SARS-CoV-2 should involve an independent audit of the WIV P4 laboratories and the laboratories of their close collaborators. Such an investigation should have taken place long ago and should not be delayed any further.

We also note that in the publication of the chimeric virus SHC015-MA15 in 2015, the attribution of funding of Zhengli Shi by the NIAID was initially left out. It was reinstated in the publication in 2016 in a corrigendum, perhaps after the meeting in January 2016 to reinstate NIH funding for gain-of-function research on viruses. This is an unusual scientific behavior, which needs an explanation for.

What is not thoroughly described in this report is the various evidence indicating that several coronaviruses recently published (RaTG13[18], RmYN02[30], and several pangolin coronaviruses[27-29,31]) are highly suspicious and likely fraudulent. These fabrications would serve no purpose other than to deceive the scientific community and the general public so that the true identity of SARS-CoV-2 is hidden. Although exclusion of details of such evidence does not alter the conclusion of the current report, we do believe that these details would provide additional support for our contention that SARS-CoV-2 is a laboratory-enhanced virus and a product of gain-of-function research. A follow-up report focusing on such additional evidence is now being prepared and will be submitted shortly.

**Acknowledgements**

We would like to thank Daoyu Zhang for sharing with us the findings of mutations in the E proteins in different sub-groups of β coronaviruses. We also thank all the anonymous scientists and other individuals, who have contributed in uncovering various facts associated with the origin of SARS-CoV-2.

*Added on July 17th, 2021:* We thank Dr. Jie Guan for helpful discussions, creating Figure 1, and proofreading the original manuscript when it was published in September 2020.

**References:**

1. Zhan, S.H., Deverman, B.E. & Chan, Y.A. SARS-CoV-2 is well adapted for humans. What does this mean for re-emergence? *bioRxiv*, https://doi.org/10.1101/2020.05.01.073262 (2020).
2. Mou, H. et al. Mutations from bat ACE2 orthologs markedly enhance ACE2-Fc neutralization of SARS-CoV-2. *bioRxiv*, https://doi.org/10.1101/2020.06.29.178459 (2020).
3. Piplani, S., Singh, P.K., Winkler, D.A. & Petrovsky, N. In silico comparison of spike protein-ACE2 binding affinities across species; significance for the possible origin of the SARS-CoV-2 virus. *arXiv*, arXiv:2005.06199 (2020).

4. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. & Garry, R.F. The proximal origin of SARS-CoV-2. *Nat Med* **26**, 450-452 (2020).
5. Maiti, A.K. On The Origin of SARS-CoV-2 Virus. *Preprint (authorea.com)*, DOI: 10.22541/au.159355977.76503625 (2020).
6. Lin, X. & Chen, S. Major Concerns on the Identification of Bat Coronavirus Strain RaTG13 and Quality of Related Nature Paper. *Preprints*, 2020060044 (2020).
7. Bengston, D. All journal articles evaluating the origin or epidemiology of SARS-CoV-2 that utilize the RaTG13 bat strain genomics are potentially flawed and should be retracted. *OSFPreprints*, DOI: 10.31219/osf.io/wy89d (2020).
8. Segreto, R. & Deigin, Y. Is considering a genetic-manipulation origin for SARS-CoV-2 a conspiracy theory that must be censored? *Preprint (Researchgate)* DOI: 10.13140/RG.2.2.31358.13129/1 (2020).
9. Rahalkar, M.C. & Bahulikar, R.A. Understanding the Origin of 'BatCoVRaTG13', a Virus Closest to SARS-CoV-2. *Preprints*, 2020050322 (2020).
10. Robinson, C. Was the COVID-19 virus genetically engineered? (https://gmwatch.org/en/news/latest-news/19383, 2020).
11. Robinson, C. Another expert challenges assertions that SARS-CoV-2 was not genetically engineered. (https://gmwatch.org/en/news/latest-news/19383, 2020).
12. Sørensen, B., Dalgleish, A. & Susrud, A. The Evidence which Suggests that This Is No Naturally Evolved Virus. *Preprint*, https://www.minervanett.no/files/2020/07/13/TheEvidenceNoNaturalEvol.pdf (2020).
13. Zhang, B. SARS-CoV-2 Could Come from a Lab - A Critique of "The Proximal Origin of SARS-CoV-2" Published in Nature Medicine. (https://www.linkedin.com/pulse/sars-cov-2-could-come-from-lab-critique-proximal-origin-billy-zhang?articleId=6651628681431175168#comments-6651628681431175168&trk=public_profile_article_view, 2020).
14. Sirotkin, K. & Sirotkin, D. Might SARS CoV 2 Have Arisen via Serial Passage through an Animal Host or Cell Culture? *BioEssays*, https://doi.org/10.1002/bies.202000091 (2020).
15. Seyran, M. et al. Questions concerning the proximal origin of SARS-CoV-2. *J Med Virol* (2020).
16. China Honors Ian Lipkin. (https://www.publichealth.columbia.edu/public-health-now/news/china-honors-ian-lipkin, 2020).
17. Holmes, E. Academic CV. (https://www.sydney.edu.au/AcademicProfiles/profile/resource?urlid=edward.holmes&type=cv, 2020).
18. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
19. Rahalkar, M. & Bahulikar, R. The Abnormal Nature of the Fecal Swab Sample used for NGS Analysis of RaTG13 Genome Sequence Imposes a Question on the Correctness of the RaTG13 Sequence. *Preprints.org*, 2020080205 (2020).
20. Singla, M., Ahmad, S., Gupta, C. & Sethi, T. De-novo Assembly of RaTG13 Genome Reveals Inconsistencies Further Obscuring SARS-CoV-2 Origins. *Preprints*, 2020080595 (doi: 10.20944/preprints202008.0595.v1) (2020).
21. Zhang, D. Anomalies in BatCoV/RaTG13 sequencing and provenance. *Preprint (zenodo.org)*, https://zenodo.org/record/3987503#.Xz9GzC-z3GI (2020).
22. Robinson, C. Journals censor lab origin theory for SARS-CoV-2. (https://www.gmwatch.org/en/news/latest-news/19475-journals-censor-lab-origin-theory-for-sars-cov-2, 2020).
23. Scientific evidence and logic behind the claim that the Wuhan coronavirus is man-made. *https://nerdhaspower.weebly.com* (2020).
24. Zhang, Y. et al. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. *bioRxiv*, https://doi.org/10.1101/2020.05.24.111823 (2020).
25. Muth, D. et al. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep* **8**, 15177 (2018).
26. Schoeman, D. & Fielding, B.C. Coronavirus envelope protein: current knowledge. *Virol J* **16**, 69 (2019).
27. Lam, T.T. et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* (2020).
28. Liu, P. et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog* **16**, e1008421 (2020).

29. Xiao, K. et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* (2020).
30. Zhou, H. et al. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr Biol* **30**, 2196-2203 e3 (2020).
31. Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr Biol* **30**, 1578 (2020).
32. Yang, X.L. et al. Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus. *J Virol* **90**, 3253-6 (2015).
33. Hu, D. et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect* **7**, 154 (2018).
34. Wang, Y. Preliminary investigation of viruses carried by bats on the southeast coastal area (东南沿海地区蝙蝠携带病毒的初步调查研究). *Master Thesis* (2017).
35. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269 (2020).
36. Lab That First Shared Novel Coronavirus Genome Still Shut Down by Chinese Government. *Global Biodefense*, https://globalbiodefense.com/headlines/chinese-lab-that-first-shared-novel-coronavirus-genome-shut-down/ (2020).
37. Song, W., Gui, M., Wang, X. & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* **14**, e1007236 (2018).
38. Li, F., Li, W., Farzan, M. & Harrison, S.C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864-8 (2005).
39. Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* (2020).
40. Hassanin, A. The SARS-CoV-2-like virus found in captive pangolins from Guangdong should be better sequenced. *bioRxiv*, https://doi.org/10.1101/2020.05.07.077016 (2020).
41. Zhang, D. The Pan-SL-CoV/GD sequences may be from contamination. *Preprint (zenodo.org)*, DOI: 10.5281/zenodo.3885333 (2020).
42. Chan, Y.A. & Zhan, S.H. Single source of pangolin CoVs with a near identical Spike RBD to SARS-CoV-2. *bioRxiv*, https://doi.org/10.1101/2020.07.07.184374 (2020).
43. Lee, J. et al. No evidence of coronaviruses or other potentially zoonotic viruses in Sunda pangolins (Manis javanica) entering the wildlife trade via Malaysia. *bioRxiv*, https://doi.org/10.1101/2020.06.19.158717 (2020).
44. Becker, M.M. et al. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc Natl Acad Sci U S A* **105**, 19944-9 (2008).
45. Menachery, V.D. et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* **21**, 1508-13 (2015).
46. Menachery, V.D. et al. SARS-like WIV1-CoV poised for human emergence. *Proc Natl Acad Sci U S A* **113**, 3048-53 (2016).
47. Ren, W. et al. Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *J Virol* **82**, 1899-907 (2008).
48. Li, X. et al. Emergence of SARS-CoV-2 through Recombination and Strong Purifying Selection. *bioRxiv* (2020).
49. Hou, Y. et al. Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. *Arch Virol* **155**, 1563-9 (2010).
50. Yang, Y. et al. Two Mutations Were Critical for Bat-to-Human Transmission of Middle East Respiratory Syndrome Coronavirus. *J Virol* **89**, 9119-23 (2015).
51. Luo, C.M. et al. Discovery of Novel Bat Coronaviruses in South China That Use the Same Receptor as Middle East Respiratory Syndrome Coronavirus. *J Virol* **92**(2018).
52. Cui, J., Li, F. & Shi, Z.L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* **17**, 181-192 (2019).
53. Wan, Y. et al. Molecular Mechanism for Antibody-Dependent Enhancement of Coronavirus Entry. *J Virol* **94**(2020).
54. Li, F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J Virol* **89**, 1954-64 (2015).

55. Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* **3**, 237-261 (2016).
56. Shang, J. et al. Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* **117**, 11727-11734 (2020).
57. Hoffmann, M., Kleine-Weber, H. & Pohlmann, S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell* **78**, 779-784 e5 (2020).
58. Coutard, B. et al. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* **176**, 104742 (2020).
59. Claas, E.C. et al. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet* **351**, 472-7 (1998).
60. Watanabe, R. et al. Entry from the cell surface of severe acute respiratory syndrome coronavirus with cleaved S protein as revealed by pseudotype virus bearing cleaved S protein. *J Virol* **82**, 11985-91 (2008).
61. Belouzard, S., Chu, V.C. & Whittaker, G.R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A* **106**, 5871-6 (2009).
62. Kido, H. et al. Role of host cellular proteases in the pathogenesis of influenza and influenza-induced multiple organ failure. *Biochim Biophys Acta* **1824**, 186-94 (2012).
63. Sun, X., Tse, L.V., Ferguson, A.D. & Whittaker, G.R. Modifications to the hemagglutinin cleavage site control the virulence of a neurotropic H1N1 influenza virus. *J Virol* **84**, 8683-90 (2010).
64. Cheng, J. et al. The S2 Subunit of QX-type Infectious Bronchitis Coronavirus Spike Protein Is an Essential Determinant of Neurotropism. *Viruses* **11**(2019).
65. Ito, T. et al. Generation of a highly pathogenic avian influenza A virus from an avirulent field isolate by passaging in chickens. *J Virol* **75**, 4439-43 (2001).
66. Canrong Wu, Y.Y., Yang Liu, Peng Zhang, Yali Wang, Hua Li, Qiqi Wang, Yang Xu, Mingxue Li, Mengzhu Zheng, Lixia Chen. Furin, a potential therapeutic target for COVID-19. *Preprint (chinaXiv)*, http://www.chinaxiv.org/abs/202002.00062 (2020).
67. Zeng, L.P. et al. Bat Severe Acute Respiratory Syndrome-Like Coronavirus WIV1 Encodes an Extra Accessory Protein, ORFX, Involved in Modulation of the Host Immune Response. *J Virol* **90**, 6573-6582 (2016).
68. Lau, S.Y. et al. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerg Microbes Infect* **9**, 837-842 (2020).
69. Liu, Z. et al. Identification of common deletions in the spike protein of SARS-CoV-2. *J Virol* (2020).
70. Ge, X.Y. et al. Detection of alpha- and betacoronaviruses in rodents from Yunnan, China. *Virol J* **14**, 98 (2017).
71. Guan, Y. et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276-8 (2003).
72. Ge, X.Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535-8 (2013).
73. Lau, S.K. et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* **102**, 14040-5 (2005).
74. Kam, Y.W. et al. Antibodies against trimeric S glycoprotein protect hamsters against SARS-CoV challenge despite their capacity to mediate FcgammaRII-dependent entry into B cells in vitro. *Vaccine* **25**, 729-40 (2007).
75. Chan, J.F. et al. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin Microbiol Rev* **28**, 465-522 (2015).
76. Zhou, J., Chu, H., Chan, J.F. & Yuen, K.Y. Middle East respiratory syndrome coronavirus infection: virus-host cell interactions and implications on pathogenesis. *Virol J* **12**, 218 (2015).
77. Yeung, M.L. et al. MERS coronavirus induces apoptosis in kidney and lung by upregulating Smad7 and FGF2. *Nat Microbiol* **1**, 16004 (2016).
78. Chu, D.K.W. et al. MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. *Proc Natl Acad Sci U S A* **115**, 3144-3149 (2018).
79. Ommeh, S. et al. Genetic Evidence of Middle East Respiratory Syndrome Coronavirus (MERS-Cov) and Widespread Seroprevalence among Camels in Kenya. *Virol Sin* **33**, 484-492 (2018).
80. Sia, S.F. et al. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. *Nature* (2020).

81. Ren, W. et al. Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. *J Gen Virol* **87**, 3355-9 (2006).
82. Yuan, J. et al. Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans. *J Gen Virol* **91**, 1058-62 (2010).
83. Ge, X.Y. et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Virol Sin* **31**, 31-40 (2016).
84. Hu, B. et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog* **13**, e1006698 (2017).
85. Luo, Y. et al. Longitudinal Surveillance of Betacoronaviruses in Fruit Bats in Yunnan Province, China During 2009-2016. *Virol Sin* **33**, 87-95 (2018).
86. Kuo, L., Godeke, G.J., Raamsman, M.J., Masters, P.S. & Rottier, P.J. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J Virol* **74**, 1393-406 (2000).
87. Drexler, J.F. et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J Virol* **84**, 11336-49 (2010).
88. Agnihothram, S. et al. A mouse model for Betacoronavirus subgroup 2c using a bat coronavirus strain HKU5 variant. *mBio* **5**, e00047-14 (2014).
89. Johnson, B.A., Graham, R.L. & Menachery, V.D. Viral metagenomics, protein structure, and reverse genetics: Key strategies for investigating coronaviruses. *Virology* **517**, 30-37 (2018).
90. Wu, K., Peng, G., Wilken, M., Geraghty, R.J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J Biol Chem* **287**, 8904-11 (2012).
91. Follis, K.E., York, J. & Nunberg, J.H. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology* **350**, 358-69 (2006).
92. Yount, B., Denison, M.R., Weiss, S.R. & Baric, R.S. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J Virol* **76**, 11065-78 (2002).
93. Yount, B. et al. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A* **100**, 12995-3000 (2003).
94. Almazan, F. et al. Construction of a severe acute respiratory syndrome coronavirus infectious cDNA clone and a replicon to study coronavirus RNA synthesis. *J Virol* **80**, 10900-6 (2006).
95. Scobey, T. et al. Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A* **110**, 16157-62 (2013).
96. Almazan, F., Marquez-Jurado, S., Nogales, A. & Enjuanes, L. Engineering infectious cDNAs of coronavirus as bacterial artificial chromosomes. *Methods Mol Biol* **1282**, 135-52 (2015).
97. Thao, T.T.N. et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* (2020).
98. Oldfield, L.M. et al. Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proc Natl Acad Sci U S A* **114**, E8885-E8894 (2017).
99. Vashee, S. et al. Cloning, Assembly, and Modification of the Primary Human Cytomegalovirus Isolate Toledo by Yeast-Based Transformation-Associated Recombination. *mSphere* **2**(2017).
100. Xie, X. et al. An Infectious cDNA Clone of SARS-CoV-2. *Cell Host Microbe* **27**, 841-848 e3 (2020).
101. Roberts, A. et al. A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mice. *PLoS Pathog* **3**, e5 (2007).
102. Roberts, A. et al. Animal models and vaccines for SARS-CoV infection. *Virus Res* **133**, 20-32 (2008).
103. Takayama, K. In Vitro and Animal Models for SARS-CoV-2 research. *Trends Pharmacol Sci* **41**, 513-517 (2020).
104. Wang, Q. hACE2 Transgenic Mouse Model For Coronavirus (COVID-19) Research. *The Jackson Laboratory Research Highlight*, https://www.jax.org/news-and-insights/2020/february/introducing-mouse-model-for-corona-virus# (2020).
105. Zhang, L. et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv*, https://doi.org/10.1101/2020.06.12.148726 (2020).
106. Yurkovetskiy, L. et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *bioRxiv*, https://doi.org/10.1101/2020.07.04.187757 (2020).

107. Korber, B. et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827 e19 (2020).
108. Plante, J.A. et al. Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. *bioRxiv*, https://doi.org/10.1101/2020.09.01.278689 (2020).
109. Poon, L.L. et al. Recurrent mutations associated with isolation and passage of SARS coronavirus in cells from non-human primates. *J Med Virol* **76**, 435-40 (2005).
110. Pervushin, K. et al. Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoS Pathog* **5**, e1000511 (2009).
111. Nieto-Torres, J.L. et al. Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathog* **10**, e1004077 (2014).